

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-306758

(43)Date of publication of application : 21.11.1995

(51)Int.Cl. G06F 3/06  
G11B 20/18  
G11B 20/18  
G11B 20/18  
G11B 20/18  
G11B 20/18

(21)Application number : 06-100454

(71)Applicant : HITACHI LTD

(22)Date of filing : 16.05.1994

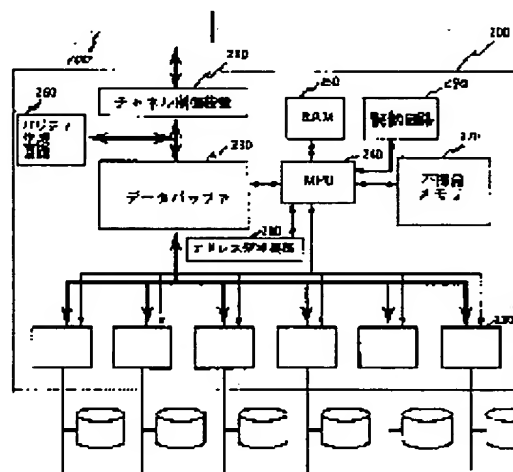
(72)Inventor : MATSUMOTO YOSHIKO  
TAKEUCHI HISAHARU  
HONMA HISAO  
SATO TAKAO

## (54) DISK ARRAY DEVICE AND ITS CONTROL METHOD

## (57)Abstract:

PURPOSE: To alter the constitution of a data file and to improve the reliability or performance of the whole device by making the redundancy of data variable during an on-line connection with a host device or turning OFF the power source of this device.

CONSTITUTION: An MPU 240 executes commands in a RAM 250 while decoding them sequentially to control the whole disk drive controller 200. At a user's request to vary the reliability of an optical logical data file, the MPU 240 of a disk drive controller 200 sets a flag (in-redundancy-variation information) indicating that the redundancy of redundancy variation information in the RAM 250 is being varied and requested redundancy (variation requested redundancy). Therefore, the number of redundant data blocks can be varied even while the disk array device is connected to the host on an on-line basis or without stopping the disk array device; and the redundancy of the logical data file is varied to flexibly comply with a request for the reliability of user data.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-306758

(43) 公開日 平成7年(1995)11月21日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	5 4 0			
G 1 1 B 20/18	5 2 0 E	8940-5D		
	5 3 2 B	8940-5D		
	5 7 0 Z	8940-5D		
	5 7 2 F	8940-5D		

審査請求 未請求 請求項の数 5 O L (全 16 頁) 最終頁に続く

(21) 出願番号 特願平6-100454

(22) 出願日 平成6年(1994)5月16日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 松本 佳子

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 竹内 久治

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 本間 久雄

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 弁理士 小川 勝男

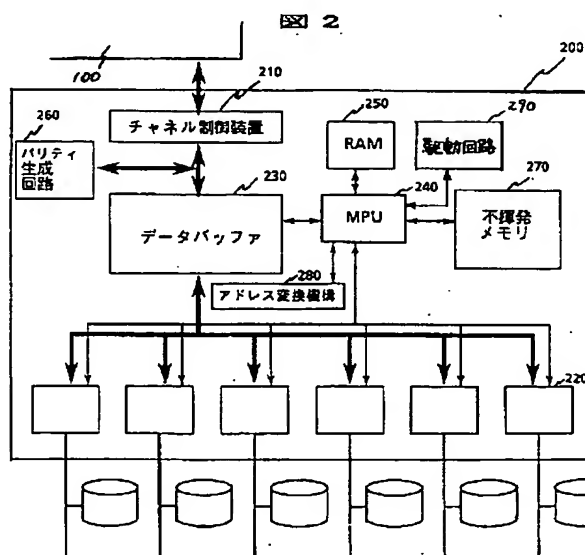
最終頁に続く

(54) 【発明の名称】 ディスクアレイ装置及びその制御方法

(57) 【要約】

【目的】 ディスクアレイ装置の動作中にデータの冗長度を変更させること、冗長度を変更させながら上位装置からの I/O 処理を実行可能とすること、冗長度の変更途中でディスクアレイ装置の電源を遮断した後、電源を投入すると当該変更処理を継続して実行すること。

【構成】 1つ又は複数の冗長度のパリティデータを生成するパリティ生成回路、ホストからの書き込み要求データ、物理ドライブからのデータブロック若しくはパリティブロック、又は、上記パリティ生成回路から生成された1つ又は複数のパリティブロックを格納するデータバッファ、少なくとも、指定された論理データファイルの RAID 構成を管理する RAID 管理情報、操作者により要求された冗長度変更情報及び物理ドライブの用途を示すデータ識別子の情報を格納する不揮発メモリ、及び、論理データの格納アドレス及び前記の RAID 管理情報から、格納対象物理ドライブのアドレス及び物理ドライブデータのアドレスを算出するアドレス変換機構を有するディスクアレイ装置。



## 【特許請求の範囲】

【請求項 1】上位装置とオンライン接続中に又は自身の電源を遮断することなくデータの冗長さの変更が可能なディスクアレイ装置。

【請求項 2】請求項 1 のディスクアレイ装置において、更に、前記冗長さを変更させながら前記上位装置から自身の論理データファイルに対する入出力処理を実行するディスクアレイ装置。

【請求項 3】請求項 1 のディスクアレイ装置において、更に、前記冗長さの変更の途中で、自身の電源を遮断した後、当該電源を投入すると、前記前記冗長さの変更を引き続き実行するディスクアレイ装置。

【請求項 4】請求項 1 のディスクアレイ装置において、更に、冗長さの変更の途中であることを表示するディスクアレイ装置。

【請求項 5】1 つ又は複数の冗長さのパリティデータを生成するパリティ生成回路、  
ホストからの書き込み要求データ、物理ドライブからのデータブロック若しくはパリティブロック、又は、上記パリティ生成回路から生成された 1 つ又は複数のパリティブロックを格納するデータバッファ、  
少なくとも、指定された論理データファイルの RAID 構成を管理する RAID 管理情報、操作者により要求された冗長さ変更情報及び物理ドライブの用途を示すデータ識別子の情報を格納する不揮発メモリ、及び、  
論理データの格納アドレス及び前記の RAID 管理情報から、格納対象物理ドライブのアドレス及び物理ドライブデータのアドレスを算出するアドレス変換機構を有するディスクアレイ装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、論理データファイルを複数のデータに分配して格納するディスクアレイ装置に係り、特に、装置（ディスクアレイサブシステム）の冗長データ格納ドライブの変更方法に関する。

## 【0002】

【従来の技術】従来の装置は、特開昭 62-24481 号公報に記載のように、上位装置からの論理データを複数のデータブロックに分割して、並列に複数の記憶媒体にアクセスすることにより、所定の性能を実現している。また並列にアクセスする単位毎にパリティデータブロックを作成し、記憶媒体のパリティデータ格納専用エリアに対し、そのデータブロックをデータの書き込みと同時に並列に書き込む。そしてデータの読み出し時に、読み出しエラーが起こったときは、障害データブロック以外のデータブロックとパリティデータブロックとを読み出し、これらのデータブロックより、障害データブロックを復元し、上位装置に転送することにより、高信頼性を維持している。

## 【0003】

【発明が解決しようとする課題】従来技術では、システムが固定的であり、ユーザの様々な要求に答えられないといった問題があった。つまり従来技術では、一度ファイルシステムを導入すると、このシステムは同一の信頼性しかない。更に高い信頼性が必要となった場合は、システム全体又は一部分を停止し（電源遮断）、ハードウェアやソフトウェアの交換が必要であった。

【0004】例えば、ある論理データファイルを冗長さ 1、分割数を 4 としてディスクアレイ装置に割当て、システムを導入する。つまり論理データファイル内の任意の論理データは、4 つの物理データブロックに分割され、かつ、1 つのパリティブロックと共に、記憶媒体に格納される。こうしておけば、この論理データファイルに対するアクセスにおいて、4 つの物理データブロック内の任意の 1 つの物理データブロックに障害が発生したとき、障害データブロック以外のデータブロックとパリティブロックを読み出し、パリティ生成回路にて障害データブロックを復元することにより、論理データを復元することができる。但し、この RAID 構成は冗長度が 1 のため、任意の論理データを構成する 4 つの物理データブロックのうち、2 つ以上の物理データブロックに障害が発生すると、データの復元は不可能である。ここで RAID とは、Redundant Arrays of Inexpensive Disks なるディスクアレイ装置の構成をいい、便宜上、クラス 1 からクラス 5 に分類されている。

【0005】この論理データファイルの信頼性を更に高めたいという要求がユーザからあったとき、この RAID 構成では、例えば、冗長度を 2 に増加する必要がある。このとき、従来技術ではディスクアレイ装置とホストとのオンライン中に、論理データファイルの構成を変更することが不可能なため、ディスクアレイ装置をいったん停止し、電源を遮断した後、電源を再投入して、システムの導入をし直さなければならなかった。

【0006】また別の例として、上記論理データファイルが、データの分配単位が複数論理データファイルである RAID クラス 5 の構成であるとき、上位装置（ホスト）からの書き込み要求があると、更新対象データブロックとパリティデータブロックをドライブからいったん読み出し、旧データブロック、旧パリティデータブロック、及び更新データブロックから新パリティデータブロックを生成し、更新データブロックと新パリティブロックを書き込む処理が必要となる。つまりデータブロックの 1 つの障害に対しては障害ブロックの復元が可能なため信頼性は維持できるが、書き込み要求時には上記処理が必要となるため、性能は低下する（この現象を「ライトペナルティ」と呼ぶ）。

【0007】この論理データファイルにおいて、信頼性よりも性能が必要であるとユーザから要求があった場合、この RAID 構成を例えば冗長さ 0 に変更する必要

があるが、従来技術ではオンライン中に論理データファイルの構成を変更することが不可能なため、やはり、ディスクアレイ装置をいったん停止し、システムを導入し直さなければならなかった。

【0008】本発明の目的は、ディスクアレイ装置が上位装置とオンライン接続中であっても、又は、ディスクアレイ装置の電源を遮断することなく、その冗長度を変更させることを可能とし、装置全体の信頼性の変更、それにもなう性能の変更が可能なディスクアレイ装置を提供することにある。◆本発明の別の目的は、冗長度を

#### 【0009】

【課題を解決するための手段】本発明の主たる構成は、

- 1) 1つ又は複数の冗長度のパリティデータを生成するパリティ生成回路、
- 2) ホストからの書き込み要求データをいったん格納する、又は、物理ドライブからのデータブロック／パリティブロックをいったん格納し、上記パリティ生成回路から生成される1つ又は複数のパリティブロックを格納するデータバッファ、
- 3) ユーザが指定した論理データファイルのRAID構成を管理するRAID管理情報；この情報には次のものが含まれる、
  - a) 各論理データファイルの分割数／分配単位／冗長度、
  - b) 冗長データのローテート単位、
  - c) 各論理データファイルが格納されている物理ドライブアドレス、
  - 4) ユーザが要求した冗長度変更情報；この情報には次のものが含まれる、
  - d) 冗長度変更中情報、変更要求冗長度、変更終了ポイント、
  - e) 冗長度変更後の論理データファイルの格納対象物理ドライブ情報、
  - 5) 物理ドライブ情報；この情報には次のものが含まれる、
  - f) 物理ドライブに格納されている論理データファイル番号、
  - g) 物理ドライブのデータ識別子（データ／パリティ／予備）、
  - 6) 上記3)、4)、5)の情報を格納する不揮発メモリ、

7) 論理データの格納アドレス（以下、アドレスをADRと表記する）及び上記3)のRAID管理情報から、格納対象物理ドライブADR、物理ドライブデータADRを算出するアドレス変換機構、を含んでいる。

#### 【0010】

【作用】ユーザから任意の論理データファイルに対する信頼性の変更の要求があったとき、ディスクドライブ制御装置のマイクロプロセッサ（後述）は、不揮発メモリにある冗長度変更情報の冗長度を変更中であることを示すフラグ（以下、冗長度変更中情報）及びユーザの要求した冗長度（変更要求冗長度）を設定する。

【0011】ユーザの要求が冗長度の削減要求である場合には、RAID管理情報のローテート単位を参照し、ローテートされていないならば、RAID管理情報の冗長度、物理ドライブADRを変更し、処理を終了する。ここで、ローテートとは、冗長データを、ある1つの物理ドライブに固定的に格納するのではなく、全物理ドライブに対しある単位で交換しながら（回しながら）格納する動作をいう。

【0012】ローテートされているときは、冗長度変更情報の冗長度変更中情報と、要求冗長度数と変更後の物理ドライブADR情報とを設定する。そして、ファイルが格納されている物理ドライブからデータブロックとパリティブロックをデータバッファに読み込み、冗長度の削減分のパリティブロックを除いて物理ドライブへの格納位置をアドレス変換機構を用いて算出し、冗長度の削減分のパリティブロックを除いて物理ドライブへデータブロック及びパリティブロックを書き込む。

【0013】この書き込み終了時点で、冗長度変更情報の変更終了ポイントを更新する。これらの処理を当該論理データファイルすべてについて終了するまで繰り返す。終了後、RAID管理情報の冗長度、物理ドライブADRを変更し、冗長度変更情報の冗長度変更中情報をリセットする。

【0014】ユーザの要求が冗長度の増加要求である場合には、冗長度変更情報の冗長度変更中情報と要求冗長度数、変更後の物理ドライブADR情報を設定する。

【0015】ローテートされていないときには、ファイルが格納されている物理ドライブからデータブロックのみをデータバッファに読み込み、パリティ生成回路で増加分のパリティブロックを生成する。そして増加分のパリティブロックを格納対象物理ドライブへ書き込む。書き込み終了時点で、冗長度変更情報の変更終了ポイントを更新する。これらの処理を論理データファイルすべてについて終了するまで繰り返す。

【0016】ローテートされているときには、ファイルが格納されている物理ドライブからデータブロックとパリティブロックをデータバッファに読み込み、パリティ生成回路で増加分のパリティブロックを生成する。データブロックから冗長度の増加分のパリティブロックを追

加して物理ドライブへの格納位置をアドレス変換機構を用いて算出し、データブロックとパリティブロックを格納対象物理ドライブへ書き込む。書き込み終了時点で、冗長度変更情報の変更終了ポイントを更新する。これらの処理を論理データファイルすべてについて終了するまで繰り返す。

【0017】上記の変更処理が全て終了したときには、RAID管理情報の冗長度及び論理データファイルが格納されている物理ドライブADRを変更し、冗長度変更情報の冗長度変更中情報をリセットする。冗長度の削減要求に従って、冗長度を減らしたことにより空いた物理ドライブは、ユーザからの指定により予備ドライブとして使用可能とする。このとき物理ドライブ情報のデータ識別子をパリティから予備ドライブへ切り替える。

【0018】次に、これらの冗長度変更の処理中にホストからのアクセスがあった場合を説明する。◆上記の変更処理中に、ホストから読み込み要求があったときは、ディスクドライブ制御装置のマイクロプロセッサは、読み込み対象論理データファイルの冗長度変更情報の冗長度変更中情報を参照し、変更中であるため変更終了ポイントを参照し、読み込み対象論理データがすでに冗長度変更済か否かを求める。

【0019】変更前であれば、RAID管理情報に従い、対象ドライブよりデータをデータバッファに読み込みホストへ転送する。◆変更済ならば、冗長度変更情報に従い、物理ドライブの追加中であればパリティデータも含めデータバッファに読み込み、データブロックのみホストに転送する。

【0020】上記の変更処理中に、ホストから書き込み要求があったときは、ディスクドライブ制御装置のマイクロプロセッサは、書き込み対象論理データファイルの冗長度変更情報の冗長度変更中情報を参照し、変更中であるため変更終了ポイントを参照し、書き込み対象論理データがすでに冗長度変更済か否かを求める。◆変更前であれば、RAID管理情報に従い、対象ドライブへデータをデータバッファから物理ドライブへ書き込む。◆変更済ならば、冗長度変更情報に従い、物理ドライブの追加中のパリティデータも含めデータバッファから物理ドライブへ書き込む。

【0021】以上のように、ディスクアレイ装置がホストとオンライン中であっても、又は、ディスクアレイ装置を停止せずに、冗長データブロック数の変更を可能とし、当該論理データファイルの冗長度を変更させることにより、ユーザデータの信頼度に対する要求に柔軟に対応が可能である。◆また、データの分配をローテートさせている時、冗長度の変更により、データの格納位置が変わる。この時、データの信頼性を高めるために、変更中に読み込んできたデータをデータバッファに格納しておき、冗長度の変更後、格納位置が変更になったデータを再度物理ドライブから読み出し、データバッファ上に

ある以前のデータとコンペアすることにより、変更により誤ってデータを書き込む動作を防止することができる。◆また、変更のための処理は長い時間を要するが、冗長度変更情報が不揮発メモリ上にあるため、当該処理の終わりを待たずにサブシステムを停止し、電源を遮断(P/S OFF)させることができる。更に次の電源を投入しIMPL (Initial Micro Program Load) 後は、まず冗長度変更情報を参照し、変更中であれば変更終了ポイントの次から冗長度変更処理を再開することにより、保守員の介入なく自動的に再実行がおこなえるので、ディスクアレイ装置の運用性能が向上する。

【0022】

【実施例】以下、本発明の一実施例を図面を用いて説明する。◆図1は本発明を適用したディスクドライブ制御装置を含む計算機システムである。◆この計算機システムは、中央処理装置であるCPU100と、ディスクドライブ制御装置200と、ディスクドライブ装置300とから主に構成される。そしてディスクドライブ制御装置200は、CPU100からの指示に従いディスクドライブ装置300を制御している。

【0023】また、上記の冗長度変更中情報や物理ドライブ情報のデータ識別子その他ディスクドライブ制御装置200の内部情報に基づいて、その状況が操作パネル201に適宜表示される。これによりシステムが、

- 1) 冗長度の変更中であること、
- 2) 物理ドライブの属性から、該ドライブをパリティ用から予備ドライブへ又はこの逆の場合へと切替えたこと、
- 3) ホストからディスクドライブ制御装置200へアクセスがあったこと、
- 4) 電源再投入により、システムの冗長性の変更動作を再開したこと、などが操作者に明示される。

【0024】図2に、ディスクドライブ制御装置200の内部構成を示す。◆マイクロプロセッサユニット(以下MPUと称す)240は、ランダムアクセスメモリ(以下RAMと称す)250を、逐次デコードしながら実行し、ディスクドライブ制御装置200全体を制御している。駆動回路290は、操作パネル201を駆動する他、別の表示装置への出力端子も具備している(図示せず)。

【0025】チャネル制御装置210は、CPU100とのデータ転送を制御している。ドライブ制御装置220は各ドライブとのデータ転送を制御している。データバッファ230は、チャネル制御装置210とドライブ制御装置220のデータ転送時に用いられるメモリである。このメモリは揮発メモリでもよいし、不揮発メモリでもよい。ここでは揮発メモリを例に説明する。

【0026】パリティ生成回路260は、CPU100より送られてきたデータに対して冗長データを生成する機能を有し、この機能はデータの復元にも用いることが

できる。冗長データを付加する単位は、上位装置から送られてきた1論理データ単位でもよいし複数の論理データ単位に対してでもよい。また、複数の論理データは、CPU100から見た1つのデータファイル内の論理データでもよいし、複数のデータファイルの論理データでもよい。ここでは、4つの論理データに対し冗長データを付加し、6つの物理ドライブに分配し格納する方法において、冗長度変更処理を例にとりあげ説明する。

【0027】不揮発メモリ270は、MPU240がオンライン中に任意の論理データファイルに対する冗長度の変更処理を行うための情報が設定されている。この情報を不揮発メモリ270に設定することにより、冗長度の変更中でも電源を遮断(P/S OFF)することを可能とし、再立上げ時に自動的に変更処理を続行することも本発明の1つの特徴である。

【0028】アドレス変換機構280は、CPU100から指示された論理データファイル番号と、論理データアドレスから、論理データファイルのRAID構成により、ディスクドライブ装置300へアクセスするときの、物理ドライブ番号、物理データアドレスを算出するためのものである。

【0029】図3は、ディスクドライブ装置300を構成する複数の磁気ディスクドライブを示している。◆データ転送制御装置310～360は、ディスクドライブ制御装置200とデータ転送を行うためのものである。各データ転送制御装置310～360には、それぞれ4台の物理ドライブ310a～310d、320a～320d、・・・、360a～360dが接続されている。

【0030】本実施例では、これらの物理ドライブ群に4つの論理データファイル(Ea、Eb、Ec、Ed)が設定されているものとする。各論理ドライブグループは、Ea=ドライブ310a～360a、Eb=310b～360b、Ec=310c～360c、Ed=310d～360dから構成され、データ回復のグループもこれと同様の構成をとっている。

【0031】次に、本実施例におけるテーブル群を図4から図8を用いて説明する。◆図4のRAID管理情報400は、各論理データファイルEa、Eb、Ec、Ed毎の、論理データファイルがいくつの物理ドライブにデータを分割しているかを示す分割数401、1つの物理ドライブに格納するデータの単位を示す分配単位402、論理データファイルにいくつの冗長データが設定されているかを示す冗長度数403、冗長データを、格納対象物理ドライブ群に対し1つの物理ドライブに固定的に格納するのか、全物理ドライブに対し所定の単位で回しながら(ローテート)格納するのかをしめすローテート単位404、論理データファイルが格納されている物理ドライブのADR(アドレス)を示す物理ドライブADR405から構成される。

【0032】ローテート単位404が0のとき、ローテート無し、つまり冗長データは1つの固定の物理ドライブへ格納される。ローテート単位404が0以外のときには、ディスクアレイ装置はRAIDクラス5を構成し、それらの物理ドライブにローテート単位404が指定した単位ごとに回しながら、論理データ及び冗長データを分配する。

【0033】図5の変更管理テーブル500は、論理データファイルに対する冗長度の変更要求を示したものである。◆実行中FLAG501が所定の内容であることは、論理データファイルが冗長度変更処理中であることを示している。冗長度数502は変更後の冗長度数を示す。格納終了ポイント503は、当該論理データファイルの冗長度の変更が終了しているADRを示す。変更後の物理ドライブADR情報504は、変更後の物理ドライブ群を示す情報である。◆冗長度を削減する場合には相当する論理データファイルを構成する物理ドライブ群が1組余ることになる。冗長度を増加する場合には、物理ドライブ群を1組追加することになる。ADR情報504は、その1組追加になる分を含めた物理ドライブADR情報を示すことになる。

【0034】図6の物理ドライブ情報600は、各物理ドライブにどの論理データファイルが格納されているか、そのデータの属性を示している。◆論理データファイル番号601は、その物理ドライブに格納されている論理データファイル番号を示している。識別子602は物理ドライブが、論理データが割当て済のドライブか予備ドライブかを示している。尚、RAID管理情報400、変更管理テーブル500、物理ドライブ情報600は不揮発性メモリ270(図2)上に設定する。

【0035】図7に本実施例で用いるデータ形式を示す。◆1つのECCグループ(パリティ生成単位)700は、4つの論理データブロック(以下データブロックと称す)701、702、703、704からなり、更に冗長データブロック(以下パリティブロックと称す)705がパリティ生成回路260(図2)により付加される(261;1パリティ生成処理)。◆更に、冗長度を1から2へ増加すると、データブロック701、702、703、704から、705、706の2つの冗長データブロックが作成される(262;2パリティ生成処理)。この5つ又は6つのデータブロックは、並列に接続されている物理ドライブに対して、例えば、図3の物理ドライブ310a、320a、330a、340a、350a、360aに対してデータ転送される。

【0036】図8はローテート有りの場合のディスクアレイ装置RAIDクラス5の構成を示し、801のRAID構成から802へ、冗長度を1つ増加させるときの概念図を示す。この論理データファイル801のRAID構成は、分割数401=4、分配単位402=2、冗長度数403=1、ローテート単位404=2、物理ド

ライブADR405=310a/320a/330a/340a/350aである。冗長度を1つ増加させて、冗長度数502=2、変更後の物理ドライブADR情報504=310a/320a/330a/340a/350a/360aとなっている(802)。

【0037】次に本発明に係るディスクアレイ装置の通常の動作を図2を用いて説明する。◆ディスクドライブ制御装置200は、通常、CPU100からの書き込み要求があると、チャネル制御装置210により、書き込み論理データを受領し、データバッファ230にいったん格納する。MPU240は、不揮発メモリ270内のRAID管理情報400(図4)の冗長度数403に従いパリティ生成回路260にてパリティブロックを生成する。このパリティブロックはいったんデータバッファ230に格納される。MPU240は、RAID管理情報400の分割数401、分配単位402、冗長度数403、ローテート単位404、物理ドライブADR405をアドレス変換機構280へ入力する。◆これにより、MPU240は、論理データの書き込み対象物理ドライブとそのADRを認識する。出力されたADR情報に  
20 従い、MPU240は、物理ドライブにデータブロック及びパリティブロックを書き込みCPU100からの書き込み要求を終了する。

【0038】CPU100からの読み込み要求があると、MPU240は、不揮発メモリ270内のRAID管理情報400の分割数401、分配単位402、冗長度数403、ローテート単位404、物理ドライブADR405をアドレス変換機構280へ入力することにより、論理データの書き込み対象物理ドライブとそのADRを認識する。出力されたADR情報に  
30 従い、MPU240は、物理ドライブにデータブロック/パリティブロックを読み込み、データバッファ230を介しホストに転送する。MPU240は、障害ブロックがあった場合には、残りのブロックより、パリティ生成回路260にて障害ブロックを回復し、ホストに転送する。

【0039】次に、ユーザから冗長ドライブの削減指示が発生した場合のディスクアレイ装置の制御方法を図9に、増設指示が発生した場合のそれを図10に、それぞれ示すフローチャートを用いて説明する。◆削減処理において、本実施例では論理データファイルEa(図3)に対する2パリティから1パリティへの削減を例にとりて説明するが、2パリティ又は1パリティからパリティなしへの変更も同様の処理方法で可能である。

【0040】まずSTEP901(図9)にて、RAID管理情報400のローテート単位404を参照し、Eaがローテートであるか否かを判断する。ローテートが有るときはSTEP902に進み、変更管理テーブル500のEaに対する実行中FLAG501を設定し、冗長度数502に削減要求である冗長度=1を設定する。また変更後の物理ドライブADR情報504には、削減

後のEaを格納する物理ドライブADR310a/320a/330a/340a/350aを設定する。◆次にSTEP903に進み格納されている物理ドライブ群よりデータブロック/パリティブロックを読み込むため、RAID管理情報400の物理ドライブADR405より対象物理ドライブADRを求める。STEP904にてSTEP903で求めた物理ドライブ群310a/320a/330a/340a/350a/360aより、データブロック/パリティブロックを読み込みデータバッファ230に格納する(RD)。

【0041】STEP905にて、変更後の論理データファイルを格納すべく物理ドライブ群をもとめるため、変更管理テーブル500の変更後の物理ドライブADR情報504より物理ドライブADRを参照する。そのADR情報、冗長度数502、RAID管理情報400の分割数401、分配単位402をアドレス変換機構280へ入力することにより、冗長度削減後の格納ADRを算出する。◆次にSTEP905で求めたADRへ4データブロック/1パリティブロックを物理ドライブへ書き込む(WRする;STEP906)。

【0042】WRが完了した時点で、冗長度変更情報500の格納終了ポインタ503を更新する(STEP907)。

【0043】STEP904からSTEP907までの処理を、論理データファイルEa全てに対して終了するまで繰り返す。STEP908にて処理が終了したと判断されたら、STEP909へ進む。◆RAID管理情報400の冗長度数403を冗長度変更情報500の冗長度数502に変更し、RAID管理情報400の物理ドライブADR405を冗長度変更情報500の変更後の物理ドライブADR情報504に変更する。◆STEP910で冗長度変更情報500の実行中FLAG501をリセットし、削減処理は終了する。

【0044】次にSTEP901にてローテート無しと判断されたときは、STEP913に進み、RAID管理情報400の冗長度数403を要求数に変更し、RAID管理情報400の物理ドライブADR405を冗長度変更情報500の変更後の物理ドライブADR情報504に変更する。◆STEP911にて空の物理ドライブは予備として使用か否かを判断し、予備ドライブとして使用するときは、STEP912に進み、物理ドライブ情報600の物理ドライブADRのデータ識別子602に予備ドライブとして使用する旨をを設定する。こうして冗長度を削減したことにより空きとなった物理ドライブの使用方法をユーザから指定できる。

【0045】ある物理ドライブへのアクセスで障害が多発したとき、物理ドライブ情報600の識別子602を参照し、予備ドライブをサーチする。予備ドライブが存在したときには、上記障害ドライブのデータを予備ドライブへコピーし、物理ドライブ情報600の論理データ

ファイル番号601を障害ドライブから予備ドライブへコピーし、識別子602も障害ドライブから予備ドライブへコピーする。また障害ドライブに格納されていた論理データファイルのRAID管理情報400の物理ドライブADR405も障害ドライブADRから予備ドライブADRに切り替える。◆このように制御することで、削減された物理ドライブをその後オンライン中に予備ドライブとして使用することも可能である。また、削減された物理ドライブを新たな論理データファイルの格納用に割当ててもかまわない。

【0046】本実施例の特徴は、RAID管理情報400、変更管理テーブル500及び物理ドライブ情報600の制御情報を持ち、複数の冗長さのパリティブロックを生成できるパリティ生成回路260を有することにより、STEP904からSTEP907、又はSTEP913の処理を行うことにより、ディスクアレイ装置とホストとのオンライン中に任意の論理データファイルの冗長数を減少できることに有る。

【0047】次に図10のプロチャートを用いて物理ドライブを増設する場合のディスクアレイ装置の制御方法を説明する。◆増設処理において、本実施例では論理データファイルEaに対する1パリティから2パリティへの増設を例にとりて説明するが、パリティなしから1パリティ又は2パリティへの変更も同様の処理方法で可能である。

【0048】まずSTEP1001にて、変更管理テーブル500のEaに対する実行中FLAG501を設定し、冗長さ数502に増設要求である冗長さ=2を設定する。また変更後の物理ドライブADR情報504には、増設後のEaを格納する物理ドライブADR310a/320a/330a/340a/350a/360aを設定する。◆次にSTEP1002に進み格納されている物理ドライブ群よりデータブロック/パリティブロックを読み込むため、RAID管理情報400の物理ドライブADR405より対象物理ドライブADRを求める。◆STEP1003にてSTEP1002で求めた物理ドライブ群310a/320a/330a/340a/350aより、データブロック/パリティブロックを読み込みデータバッファ230(図2)に格納する(RD)。

【0049】次にパリティ生成回路260より2パリティを生成しデータバッファ230へ増設分のパリティブロックも格納する(STEP1004)。◆STEP1005にてRAID管理情報400のローテート単位404を参照し、Eaがローテートであるか否かを判断する。◆ローテートが有るとき、STEP1006に進み、変更後の論理データファイルを格納すべく物理ドライブ群をもとめるため、変更管理テーブル500の変更後の物理ドライブADR情報504より物理ドライブADRを参照する。そして、ADR情報、冗長さ数50

2、RAID管理情報400の分割数401及び分配単位402をアドレス変換機構280へ入力することにより冗長さ増設後の格納ADRを算出する。◆STEP1007にて、STEP1006で求めたアドレスADRへ4データブロック/2パリティブロックを物理ドライブへWRする。例えば、図8に示す様なRAID構成の場合、論理データ15は冗長さ1のときには物理ドライブ350aの物理ADR2に格納されている。冗長さ2になると物理ドライブ360aの物理ADR2に変更される。

【0050】STEP1005にてローテートがないときSTEP1008に進み、変更後の物理ドライブADR情報504とRAID管理情報400の物理ドライブADR405を比較することにより、追加物理ドライブをサーチし、当該追加物理ドライブへ追加パリティを書き込む(WRする)。◆STEP1007/1008のWRが完了した時点で、STEP1009に進み冗長さ変更情報500の格納終了ポインタ503を更新する。◆STEP1004からSTEP1009までの処理を論理データファイルEaの全てに対して終了するまで繰り返す。

【0051】STEP1010にて処理が終了したと判断されたら、STEP1011へ進み、RAID管理情報400の冗長さ数403を冗長さ変更情報500の冗長さ数502に変更し、RAID管理情報400の物理ドライブADR405を冗長さ変更情報500の変更後の物理ドライブADR情報504に変更する。◆STEP1012で冗長さ変更情報500の実行中FLAG501をリセットする。◆STEP1013で物理ドライブ情報600の追加物理ドライブADRのデータ識別子を予備からデータに変更し増設処理は終了する。

【0052】本実施例の特徴は、RAID管理情報400、変更管理テーブル500及び物理ドライブ情報600の制御情報を持ち、複数の冗長さのパリティブロックを生成できるパリティ生成回路260を有し、STEP1004からSTEP1009の処理を行うことにより、ディスクアレイ装置とホストとのオンライン中に任意の論理データファイルの冗長数を増加できることに有る。

【0053】次に、図9又は図10に示す変更処理を実行しているときの論理データファイルへのホストからのアクセス方法を図11のフローチャートを用いて説明する。◆CPU100(図1)からアクセス要求があると、ディスクドライブ制御装置200のMPU240(図2)は、STEP1101にて要求された論理データファイルの変更管理テーブル500の実行中FLAG501を参照する。◆STEP1102に進み、変更中か否かを判定する。もし変更中ならば、STEP1103に進み、対象論理データファイルの冗長さ変更情報500の格納終了ポインタ503を参照する。



【0054】次にSTEP1104にてアクセス対象論理データが、すでに冗長度変更済か否かを判定する。もし変更済ならば、STEP1105に進み、冗長度変更情報500の冗長度数502、変更後の物理ドライブADR情報504、RAID管理情報400の分割数401、分配単位402、ローテート単位404を参照する。◆STEP1102にて変更中でないと判断されたとき、又はSTEP1104にて終了済でないと判断されたとき、STEP1106に進む。そして、RAID管理情報400の分割数401、分配単位402、ローテート単位404物理ドライブADR405を参照する。

【0055】STEP1105又はSTEP1106からSTEP1107に進み、CPU100からの要求が読み込み要求か否かを判断する。◆読み込み処理であると判断されたとき、STEP1108に進む。そして、STEP1105の情報をアドレス変換機構280に入力し、読み込み対象論理データの物理ドライブ格納ADRを算出する。◆次にSTEP1110に進み、STEP1108の物理ドライブADRからデータブロック／パリティブロックを読み出し（RD）、データバッファ230へ格納する。◆STEP1112にてデータバッファよりパリティブロックを除いてホストへ論理データ転送する。

【0056】STEP1107にて書き込み要求であると判断したとき、STEP1106の情報をアドレス変換機構280に入力し、書き込み対象論理データの物理ドライブ格納ADRを算出する（STEP1109）。◆次にSTEP1111にて書き込み対象データをCPU100よりデータバッファ230に受領し、パリティ生成回路260にてパリティを生成し、データバッファ230に格納する。◆次にSTEP1113に進み、データバッファよりDRVへデータブロック／パリティブロックを書き込む（WRする）。

【0057】このように、本実施例によれば、RAID管理情報400、変更管理テーブル500、物理ドライブ情報600、複数のパリティブロックを生成できるパリティ生成回路260、図11のSTEP1101～1113により、論理データファイルの冗長度の変更処理を実行中にCPU100からのI/Oを受け付けることができ、ディスクアレイ装置とホストとのオンライン中の変更処理が可能となる。

【0058】また、本実施例によれば、冗長度変更処理にてデータブロックをいったんデータバッファに格納しADR変換をかけて物理ドライブに格納するので、格納し終わった後に再度物理ドライブから読み直し、データバッファ上にてデータを比較することにより、変更後のデータの信頼性を高めることもできる。

【0059】また、本実施例によれば、次の方法により電源遮断（P/S OFF）後の再立ち上げ時に自動的

に変更処理を継続することも可能である。◆即ち、再立ち上げ時に変更管理テーブル500の実行中FLAG501が設定されているかを判定する。変更中であれば、図9のSTEP903から処理を続行することにより自動的に変更処理を継続することができる。◆この方法は、RAID管理情報400変更管理テーブル500、物理ドライブ情報600が不揮発性メモリに設定されているため、変更の途中で電源を遮断（P/S OFF）されても、再立ち上げ時に情報が残っていることにより実行される。◆このように、変更処理の中断情報を常に不揮発メモリに設定しておくことにより、電源を投入（P/S ON）後、保守員の人手を介さずに変更処理を行うことができる。

【0060】念のため、以下に本発明において用いられる方法につき、特徴点を交えて整理し、列記する。

【0061】1） 1グループ内がデータ格納用ドライブと、冗長データ格納用ドライブから構成される複数の物理ドライブで構成される論理ドライブグループを、複数有するディスクドライブ装置と、前記ディスクドライブ装置と、上位装置との間に介在し、両者間における情報の転送を制御をするため、当該情報の分配又は収集を行うためのデータバッファと、一部に不揮発メモリを有し、更に分割したデータから可変数の冗長データを作成するECC生成機能を備えたディスクドライブ制御装置とからなるディスクアレイ装置において、オンライン中に操作者（ユーザ）から任意のデータファイルに対する信頼性向上の要求があったとき、データファイルに対応する論理ドライブグループが構成されているデータ格納用ドライブからいったんデータバッファに読み込み、ECC生成機能を用いて作成される冗長データ数を変更し、データバッファに作成された冗長データ群を冗長データ格納用ドライブに対してデータバッファより格納し、更に、不揮発メモリに存在する論理ドライブグループの冗長数又は格納対象物理ドライブを管理している管理情報を変更することにより、論理ドライブグループの冗長データ数を増加させ、装置の動作中（オン中）に信頼性を向上させることを特徴とする冗長データ数変更方法。

【0062】2） 上記1）のディスクアレイ装置において、オンライン中にユーザから任意のデータファイルに対する信頼性削減の要求があったとき、不揮発メモリに存在する論理ドライブグループの冗長数又は格納対象物理ドライブを管理している管理情報を変更することにより、論理ドライブグループの冗長データ数を削減させ、オン中に信頼性を削減させることを特徴とする冗長データ数変更方法。

【0063】3） 上記1）のディスクアレイ装置において、上記1）のように任意の論理ドライブグループ内の冗長データ数を削減したとき、以前に当該論理ドライブグループ内の冗長データ格納用の物理ドライブとして

割当てられていたものを、ユーザからの指定により、予備ドライブとして割当てるとき、不揮発メモリに存在する予備ドライブ管理情報を変更することにより、ドライブの障害が発生したとき、上記管理情報を参照することにより、予備ドライブである物理ドライブを認識し、当該物理ドライブを含む冗長データグループより、障害ドライブのデータを復元し、割当てた予備ドライブに格納することにより、削減した物理ドライブを予備ドライブとして使用することを特徴とする冗長ドライブ変更管理方法。

【0064】4) 上記1)のディスクアレイ装置において、上位から論理データを書き込むときは、不揮発メモリ内の当該論理ドライブグループの冗長数又は格納対象物理ドライブの管理情報を参照し、当該冗長データ数に従い、ECC生成機能を使用して冗長データを生成し、対象ドライブに格納することにより、冗長数変更後にホストから書き込むデータに対する信頼性も自動的に変更可能とすることを特徴とする冗長ドライブ変更管理方法。

【0065】5) 上記1)のディスクアレイ装置において、上位から論理データを読み込むときは、不揮発メモリ内の当該論理ドライブグループの冗長数又は格納対象物理ドライブの管理情報を参照し、当該冗長データ数と格納ドライブに従い、データ及び冗長データをデータバッファに読み込み、読み込み時に障害が発生したときは、データと共に読み込んである冗長データ群よりECC生成機能を使用してデータを復元させ、ホストに転送することにより、冗長数変更後にホストから読み込むデータに対する信頼性も自動的に変更可能とすることを特徴とする冗長ドライブ変更管理方法。

【0066】6) 上記1)における変更方法で、冗長データを物理ドライブにローテートして格納するRAID構成において、当該データバッファに格納した論理データを元にECC生成機能により新たに変更された冗長数分の冗長データを作成し、更に冗長データ数が変化したことによる論理データ及び冗長データ群の格納位置をアドレス変換機構を用いて算出し、論理データ群と冗長データ群を各々の格納対象位置の物理ドライブアドレスに格納することを特徴とする冗長ドライブ変更管理方法。

【0067】7) 上記1)のディスクアレイ装置で、冗長データを物理ドライブにローテートして格納するRAID構成において、上記2)における冗長データ数削減方法においても上記5)の変更方法により、実現することを特徴とする冗長ドライブ変更管理方法。

【0068】8) 上記1)の変更処理において、冗長データ数をECC生成機能により変更して生成し、当該冗長データを対象となる物理ドライブに格納したとき、格納が終了しているアドレスを不揮発メモリに覚えておき、前記変更処理中の論理データファイルに対してホス

トから書き込み要求があると、前記アドレスを参照し、変更以前の領域への書き込み要求であれば、変更以前の冗長数の冗長データを生成して対象ドライブへ書き込み、変更後の領域への書き込み要求であれば、変更後の冗長数の冗長データを生成して対象ドライブへ書き込むことにより、変更中のホストからのアクセスを受け付け、性能低下を防止することを特徴とした冗長ドライブ変更管理方法。

【0069】9) 上記2)の変更処理において、前記変更処理中の論理データファイルに対するホストから読み込み要求があると、前記不揮発メモリに格納されているアドレスを参照し、変更以前の領域への読み込み要求であれば、当該論理データを構成している物理データと、変更以前の冗長数の冗長データを対象ドライブより読み込み、変更後の領域への読み込み要求であれば、変更後の冗長数の冗長データを対象ドライブより読み込むことにより、変更中のホストからのアクセスを受け付け、性能低下を防止することを特徴とした冗長ドライブ変更管理方法。

【0070】10) 上記1)の外部記憶装置で、上記8)記載、上記9)記載の変更方法において、冗長数の変更の際、変更された冗長データ格納用ドライブへの冗長データの格納が終了しているポイントを不揮発メモリに覚えておくことにより、ディスクアレイ装置のP/S OFF後の再立ち上げのときに、自動的に変更処理を続行することを特徴とした冗長ドライブ変更管理方法。

【0071】11) 上記1)のディスクアレイ装置において、上記7)又は上記8)記載の変更処理をはじめる前に、変更対象となる論理データグループ内のデータをいったんデータバッファに読み込み、変更処理が終了した時点で、変更対象論理データを再格納した物理ドライブから論理データを読み出し、データバッファ上の変更前のデータと変更後のデータをコンペアすることにより、オンライン中の冗長データ変更処理のデータの信頼性を高めることを特徴とした冗長ドライブ変更管理方法。

【0072】12) 上記1)のディスクアレイ装置において、上記1)記載の冗長データ数の増加における格納対象冗長ドライブが予備ドライブとして割当てられているか、又はオンライン中に増設されることを特徴とする冗長ドライブ変更管理方法。

【0073】

【発明の効果】本発明に係るディスクアレイ装置によれば、ディスクアレイ装置の動作中にデータの冗長度を変更させることが可能であり、ディスクアレイ装置(サブシステム)を停止することなくデータファイルの構成を変更できるので、装置全体の信頼性又は性能を向上させることができる。

【0074】更に、変更処理が長時間を有しても、処理実行中での電源遮断(P/S OFF)を可能としてい

る。即ち次の電源投入（P/S ON）時には、中断した時点から保守員の介入なく自動的に変更処理の再開を可能としているため、保全性能の向上につながり、使い勝手が良くなる効果がある。

【図面の簡単な説明】

【図1】本発明の一実施例であるディスクドライブ装置を含む計算機システムの構成を示すブロック図である。

【図2】本発明の一実施例であるディスクドライブ装置の一部をより詳細に示す概念図である。

【図3】本発明の一実施例であるディスクドライブ装置を構成するディスクアレイ装置のドライブ構成を示す図である。

【図4】本発明の一実施例であるディスクアレイ装置において用いられるRAID管理情報を示す概念図である。

【図5】本発明の一実施例であるディスクアレイ装置において用いられる変更管理テーブルを示す概念図である。

【図6】本発明の一実施例であるディスクアレイ装置において用いられる物理ドライブ情報を示す概念図である。

【図7】本発明の一実施例であるディスクアレイ装置において用いられる論理データのデータ形式を示す概念図である。

【図8】本発明の一実施例であるディスクアレイ装置の冗長度の変更処理において、ローテート有り（分割数＝4／分配単位＝2／ローテート単位＝2）の場合に、冗長度を1から2へ変更するデータ格納形式の一例を示す概念図である。

【図9】本発明の一実施例であるディスクアレイ装置の制御方法を示すフローチャートである。

【図10】本発明の一実施例であるディスクアレイ装置

の制御方法を示すフローチャートである。

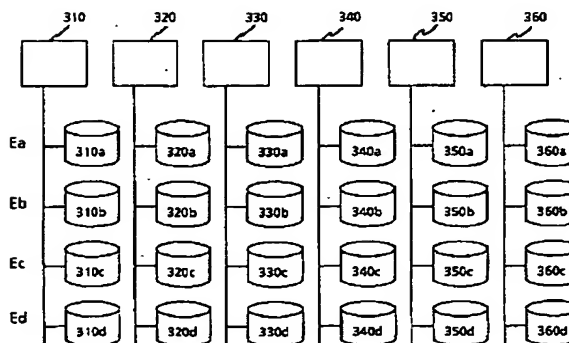
【図11】本発明の一実施例であるディスクアレイ装置の制御方法を示すフローチャートである。

【符号の説明】

100；中央処理装置（CPU）、200；ディスクドライブ制御装置、210；チャネル制御装置、220；ドライブ制御装置、230；データバッファ、240；マイクロプロセッサユニット（MPU）、250；RAM、260；パリティ生成回路、261；1パリティ生成、262；2パリティ生成、270；不揮発メモリ、280；アドレス変換機構、300；ディスクドライブ装置、310～360；データ転送制御回路、310a～360a；物理ドライブ（論理データファイルEa）、310b～360b；物理ドライブ（論理データファイルEb）、310c～360c；物理ドライブ（論理データファイルEc）、310d～360d；物理ドライブ（論理データファイルEd）、400；RAID管理情報、401；分割数、402；分配単位、403；冗長度数、404；ローテート単位、405；物理ドライブADR、500；変更管理テーブル、501；実行中FLAG、502；冗長度502、503；格納終了ポイント503、504；追加物理ドライブADR504、600；物理ドライブ情報、601；論理データファイル番号、602；データ識別子、800；ECCグループ（パリティ生成単位）、801～804；論理データブロック、805～806；パリティブロック、901～913；冗長度削減処理フロー、1001～1013；冗長度増加フロー、1100～1113；ホストアクセス処理フロー。

【図3】

図 3



【図4】

図 4

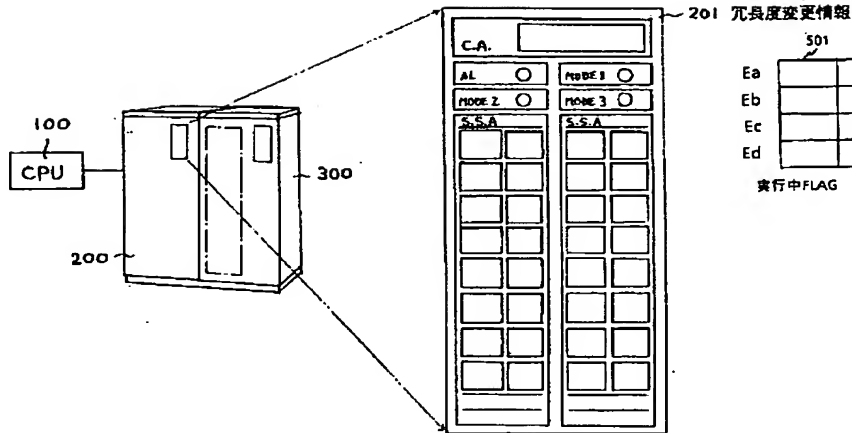
RAID管理情報

	401	402	403	404	405
Ea					
Eb					
Ec					
Ed					

分割数 分配単位 冗長度数 ローテート単位 物理ドライブADR

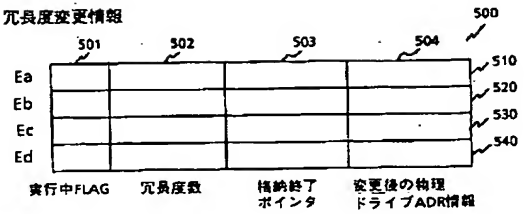
【図 1】

図 1



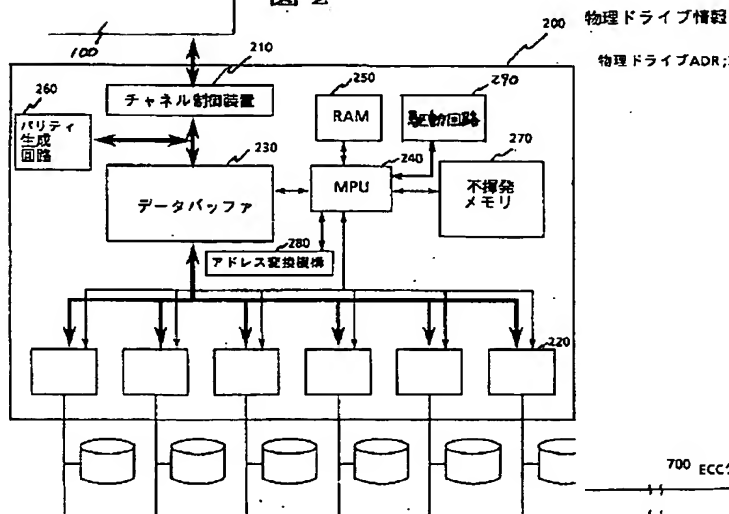
【図 5】

図 5



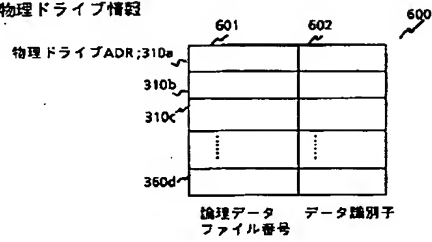
【図 2】

図 2



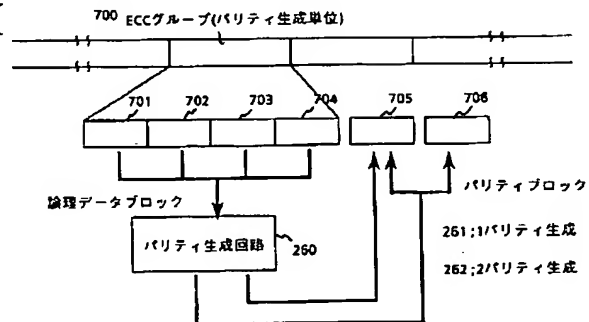
【図 6】

図 6



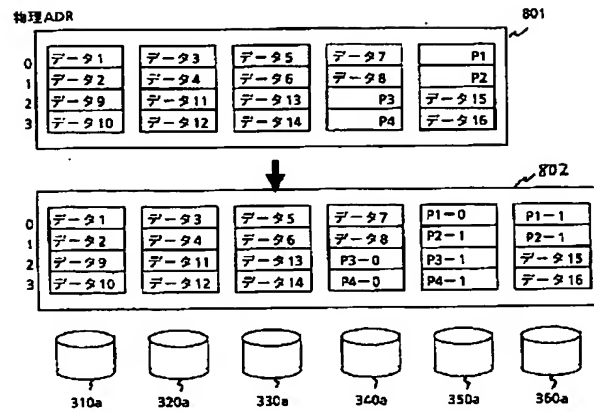
【図 7】

図 7

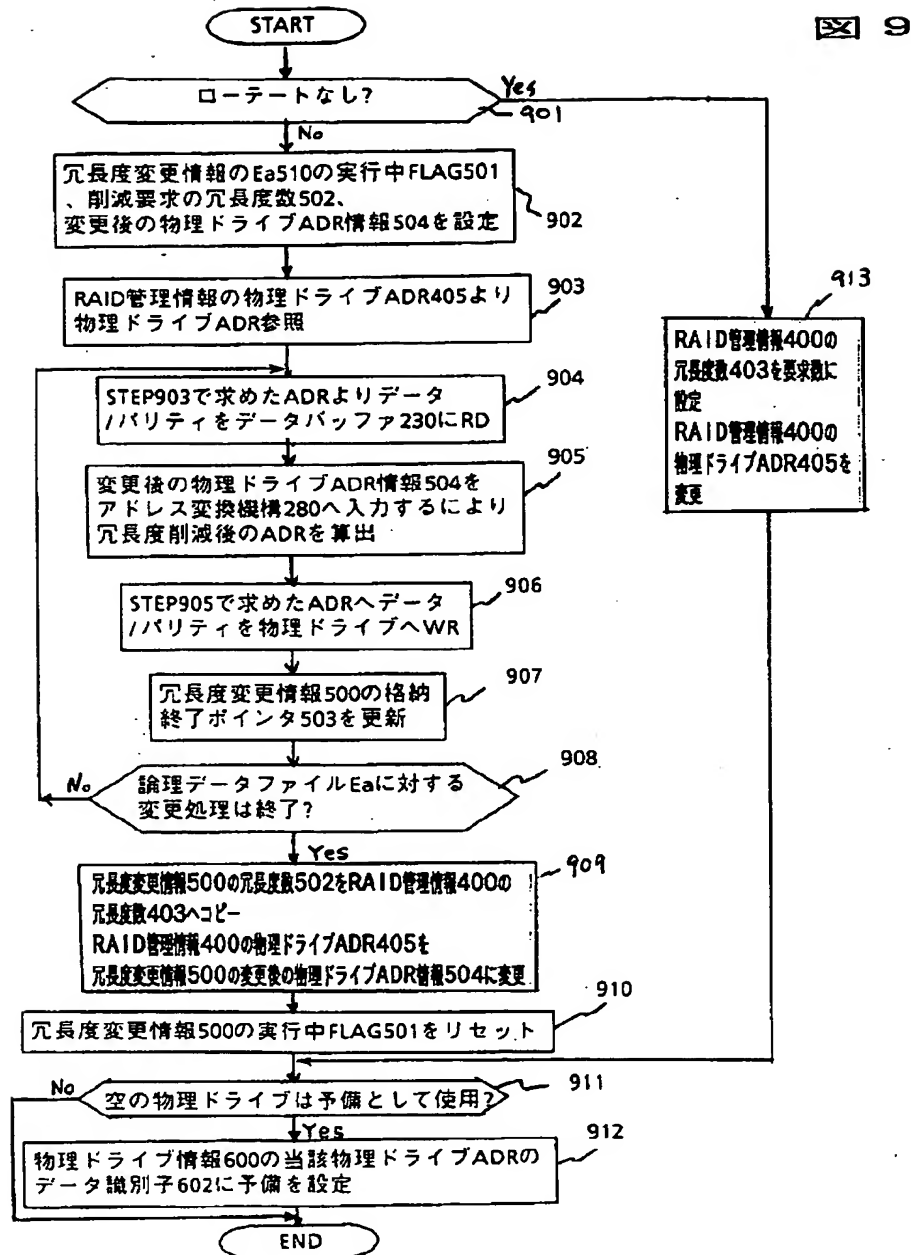


【図 8】

図 8

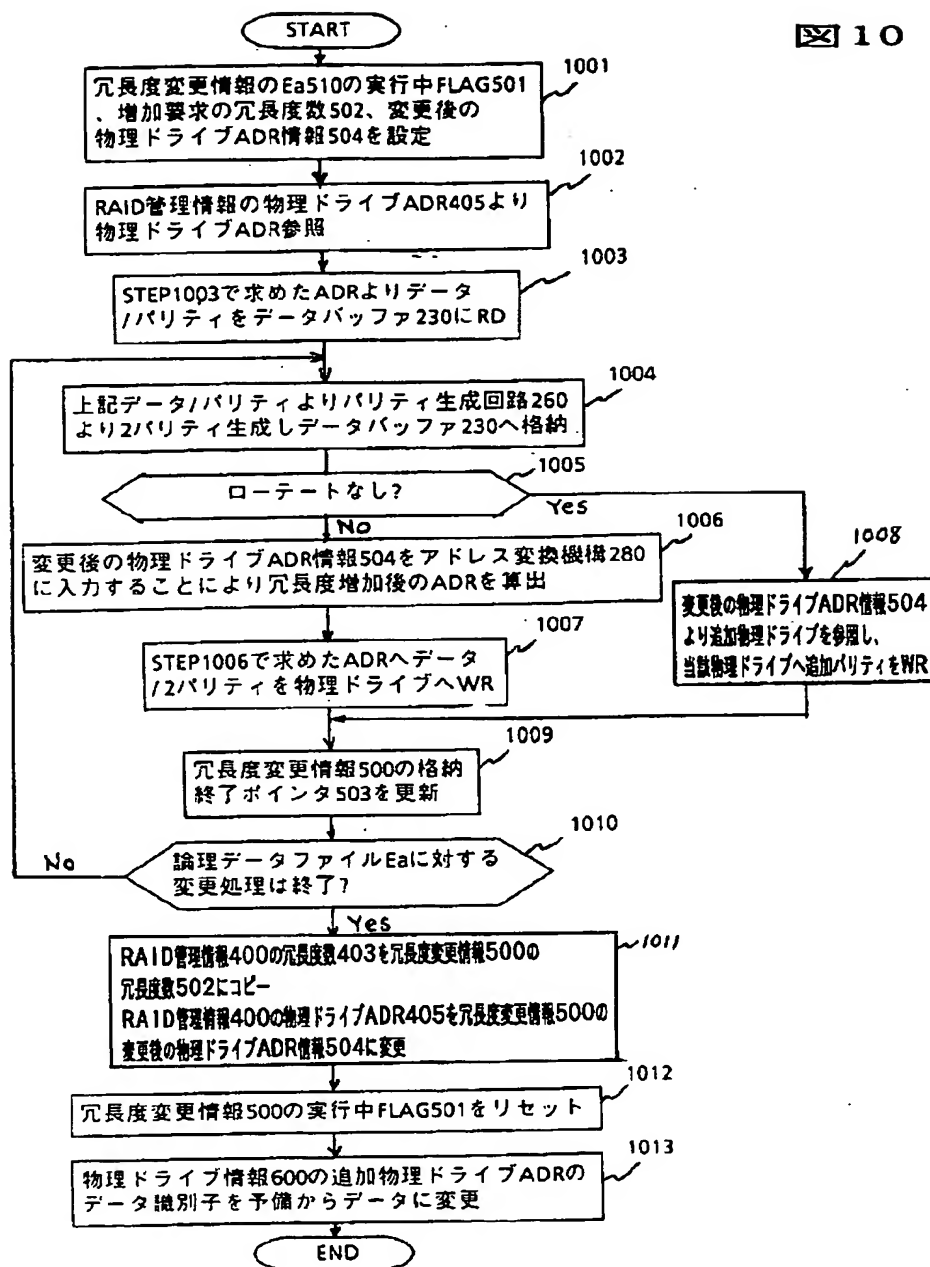


【図9】

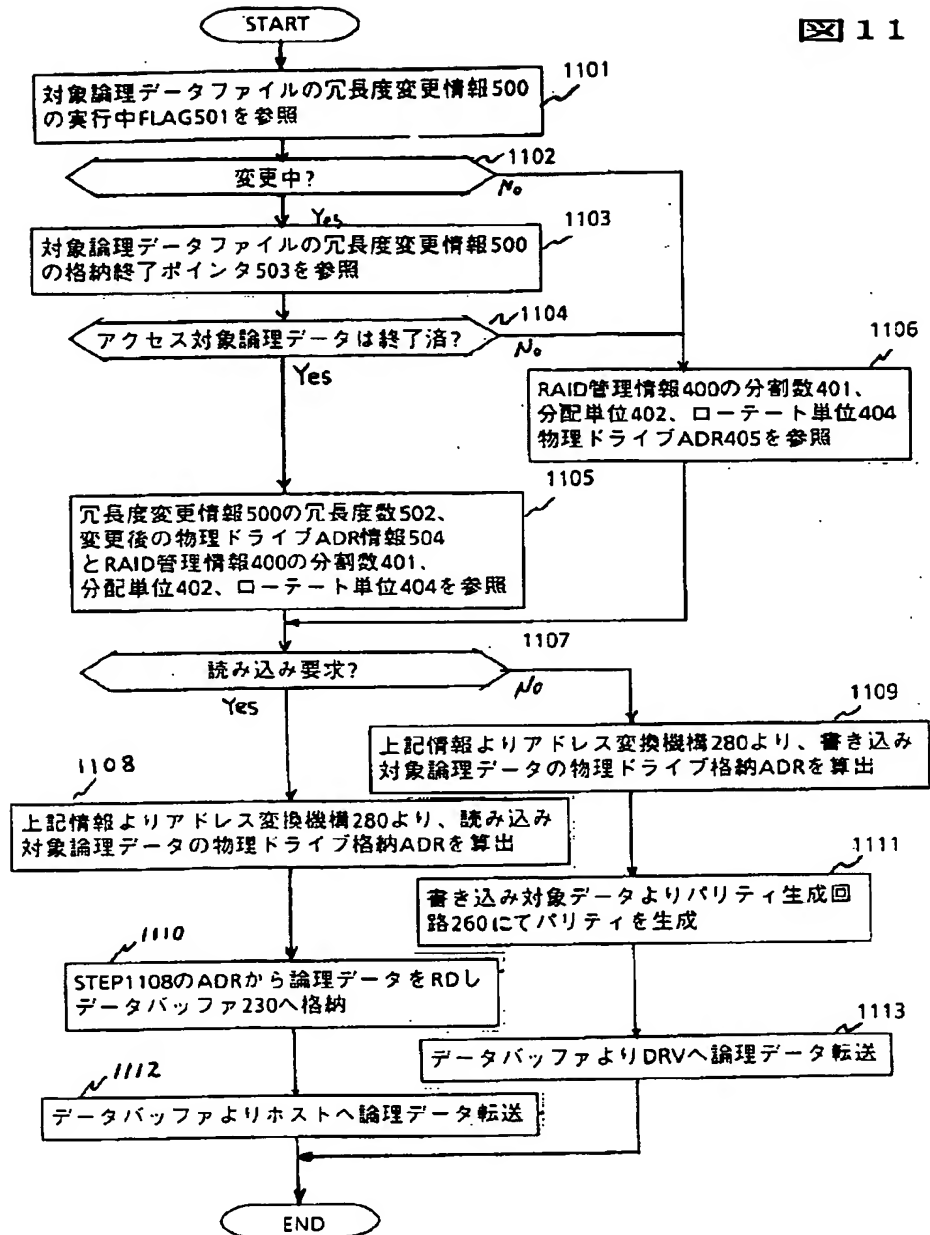


【図10】

図 10



【図11】



フロントページの続き



(72) 発明者 佐藤 孝夫  
神奈川県小田原市国府津2880番地 株式会  
社日立製作所ストレージシステム事業部内